7.5 A 0.1μm 1.8V 256Mb 66MHz Synchronous Burst PRAM

Sangbeom Kang, WooYeong Cho, Beak-Hyung Cho, Kwang-Jin Lee, Chang-Soo Lee, Hyung-Rock Oh, Byung-Gil Choi, Qi Wang, Hye-Jin Kim, Mu-Hui Park, Yu-Hwan Ro, Suyeon Kim, Du-Eung Kim, Kang-Sik Cho, Choong-Duk Ha, Youngran Kim, Ki-Sung Kim, Choong-Ryeol Hwang, Choong-Keun Kwak, Hyun-Geun Byun, Yun Sueng Shin

Samsung, Hwasung, Korea

Ongoing progress in PRAM technology since early 2000 has confirmed its great potentials as a next-generation high-density non-volatile memory [1, 2, 3]. Beside the simple cell structure with superior scalability [4], one of the most fundamental advantages of PRAM over the existing non-volatile memories is its write performance. For example, Flash memory, used primarily in mass storage applications, undergoes some inherent restrictions that PRAM doesnot have to suffer from. In Flash memory, since old data cannot be overwritten by new data, the entire block has to be frequently erased before programming new data, conflicting with its relatively low program/erase endurance limit. This requires sophisticated data handling algorithms, like garbage-collection and wear-leveling. In this paper, a 256Mb PRAM implemented in 0.1μm CMOS technology, featuring 66MHz burst-read operation, is presented. To supply sufficient write current from write-driver to the accessed cell and evaluate write throughput characteristics under the low supply voltage (1.8V), a particular parallel organization of charge-pump units is utilized.

The 256Mb PRAM is configured using global BL (GBL) architecture to make the chip area-efficient, in which sense amplifiers (S/As) and write drivers (W/Ds) are globally placed. The 256Mb PRAM consists of 4 banks. Each bank contains 16 blocks and each block is comprised of four 1Mb sub-array blocks, as shown in Fig. 7.5.1. The 1Mb sub-array is built with 1024 BLs and 1024 WLs. For the increased data transfer rates, synchronous burst-read capabilities are offered. In burst-read operation, the entire 4k cells connected to the selected WLs of a block are activated by global and local WL drivers (MWD and SWD). From each 1Mb sub-block, a single word (×16) information is read out, producing 4 word data to be latched simultaneously at 64 S/A outputs. The pre-fetched 4 word data are serially transferred from 64 S/As to the DQ driver in synchronization with clock. By pre-fetching 4 words within 62ns, it is possible to perform burst-read operation at 66MHz. By setting mode-configuration register (MCR), various choices of 'initial wait cycles' and 'burst length' are possible. In default mode, after an address access, expected data are burst out after 7 wait cycles in continuous mode. The initial access time (tIAA) is 62ns and the burst access-time valid clock to output delay (tBA) is 10ns, as confirmed by Shmoo plots in Fig. 7.5.2.

With the adoption of GBL architecture for high density, the SET or RESET write current supplied from write driver to the cell passes through parasitic resistances of global and local column-selection switches (GY and LY), and global and local BL resistances (RGBL and RLBL), as shown in Fig. 7.5.1. With an external power supply of 1.8V, the voltage level at the GST cell becomes too low to flow sufficient write current, due to the parasitic resistances. To overcome this obstacle, the power-supply level for write driver is elevated from 1.8V to VPP_WD (4 to 4.5V) by charge pumps during the write operation. In addition, to reduce the channel resistances, the voltage levels applied to the gates of MOSFETs for cell and column-selection are also boosted to VPP_X (3V) and VPP_Y (4V), respectively. To reduce the stand-by current, and to ensure device reliability by minimizing the duration of high-voltage exposure, all the pumps stop operation with their outputs discharged to V_{dd} level when not in write operation. In Fig. 7.5.3, the effects of write-pulse width on the distributions for SET and RESET resistances are shown. For reasonable SET resistance distribution, a SET-pulse width of at least 300ns is required. For RESET operation, 50ns pulse width is enough to achieve sound RESET resistance distribution and sensing margin. 500ns is chosen for both RESET and SET current pulse width.

The charge-pump system for write driver is designed to cope with high output-load currents which are in the range of several mAs, sustaining the target output voltage of 4 to 4.5V. Figure 7.5.4(a) shows the organization of WD_PUMP that consists of 8 sub-pumps (WD_PUMP1 to WD_PUMP8) connected to VPP_WD node in parallel. The basic element for each sub-pump is built with modified Dickson charge pump, with 2mA drive capability. Using 'write-mode selector', the number of enabled sub-pumps can be selected as ×2, ×4, or ×8, corresponding to the demand for current delivery, which is determined from the number of cells of parallel writing. Figure 7.5.4(b) is the measurement result of current delivering capability of a sub-pump. With one pump enabled, 4 to 4.5V output voltage is sustainable with load currents of up to ~2mA. When all pumping levels reach the target value, a flag signal (Level Flag) is triggered. The actual write operation starts only with this flag set as high, in order to prevent the improper RESET/SET write due to insufficient pumping level.

To RESET one GST cell, 600μA write current is required. In worst case, this may be close to 1mA depending on the variation of cell dimension. For simultaneous RESET operation on 16 cells (×16 internally), approximately 10mA should be supplied from WD with all the sub-pumps enabled. In this case, the current consumption of WD_PUMP is ~70mA with 35% power efficiency, as shown in the write performance characterization of Fig. 7.5.5. When this PRAM is used in a mobile system, the instantaneous amount of current that the system can tolerate may be restricted, due to noises at the power line during the pump operation. Practically, the number of cells for parallel writes needs to be reduced down to ×8, ×4, or ×2 to maintain instantaneous current consumption in an acceptable range. Instead, number of write cycles is increased by 2, 4, or 8 times, that results in increased overall write time (Fig. 7.5.5). Thus, to reduce the write time by increasing the cell-parallel-write, it is critical to develop a cell material and structure that consumes smaller amount of current for phase change. In Fig. 7.5.6, the write throughputs of the designed PRAM are summarized with those of NOR Flash memories. While the program throughput of commercially available NOR Flash products is ~0.2 MB/s, but it can be further enhanced to 1.5MB/s with improved programming algorithms[5]. The write throughput for the PRAM of the proposed design is 0.48MB/s when operated in internal ×2 mode, and can be increased to 3.3MB/s with ×16 parallel-write and decreased SET time from 500ns to 300ns, which is acceptable as determined from Fig. 7.5.3.

In summary, a 256Mb PRAM is developed, featuring 66MHz synchronous burst-read operation. Considering the current consumption, write performance is characterized at 1.8V supply. Figure 7.5.7 is a micrograph of the chip, fabricated with 0.1μm technology using 1 W-damascene and 2 Al-metal layers. The chip size is 79.2mm^2 with a unit cell size of 0.166μm^2.

References:

[1] Y. N. Hwang, et al., "Full Integration and Reliability Evaluation of Phase-Change RAM Based on 0.24μm-CMOS Technologies," *Symp. VLSI Tech.,* pp.173–174, June, 2003.

[2] W. Y. Cho, et al., "A 0.18μm 3.0-V 64-Mb Nonvolatile Phase-Transition Random Access Memory (PRAM)," *IEEE J. Solid-State Circuits,* pp. 293-300, Jan., 2005.

[3] H. R. Oh, et al.,"Enhanced Write Performance of a 64 Mb Phase-Change Random Access Memory," *ISSCC Dig. Tech. Papers,* pp. 48-49, Feb., 2005.

[4] Yun Seung Shin, "Non-Volatile Memory Technologies for Beyond 2010," *Symp. VLSI Circuits,* pp. 156-159, 2005.

[5] M. Taub, et al.,"A 90nm 512Mb 166MHz Multilevel Cell Flash Memory with 1.5 MB/s Programming," *ISSCC Dig. Tech. Papers,* pp. 54-55, Feb., 2005.

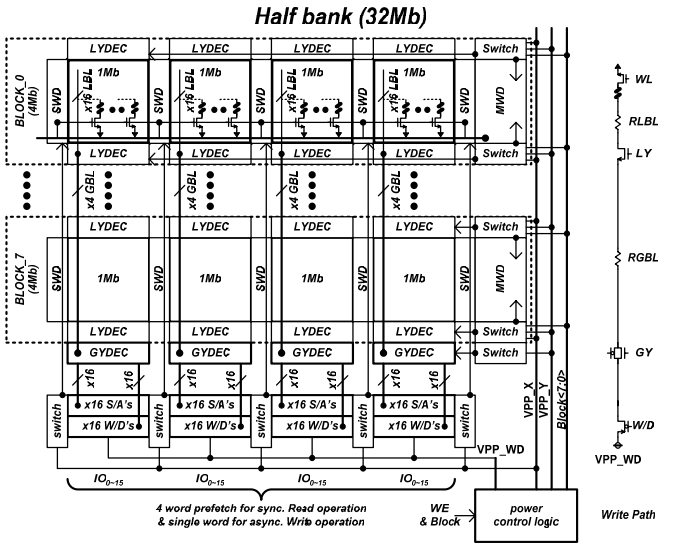

Figure 7.5.1: 32Mb half-bank architecture containing 32 1Mb sub-cores with global sense amplifiers and write drivers.

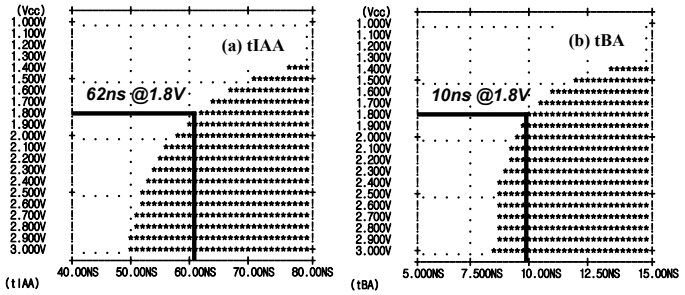

Figure 7.5.2: Shmoo plots for initial access time (tIAA) and burst access time (tBA).

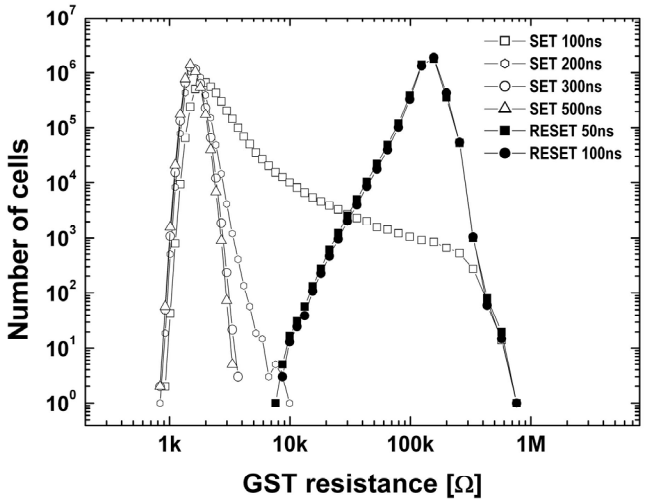

Figure 7.5.3: Distributions of GST cell resistances with varied duration of RESET/SET pulses, based on a 4Mb sample.

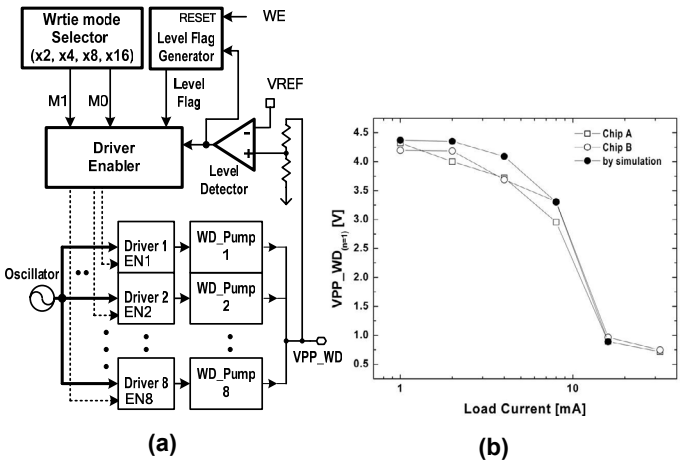

Figure 7.5.4: Charge pump organization for write driver (a), and pumping capability of one sub-pump (b).

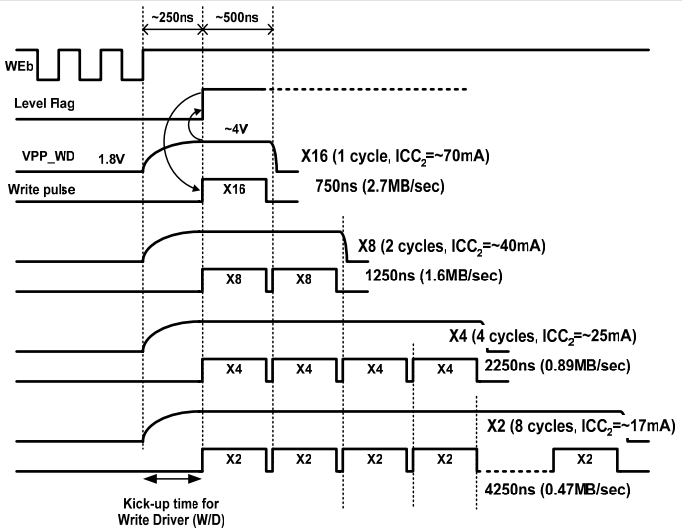

Figure 7.5.5: Write performance characterization, adjusting number of cell-parallel-write (X2 -X16).ICC2 is current consumption during write operation.

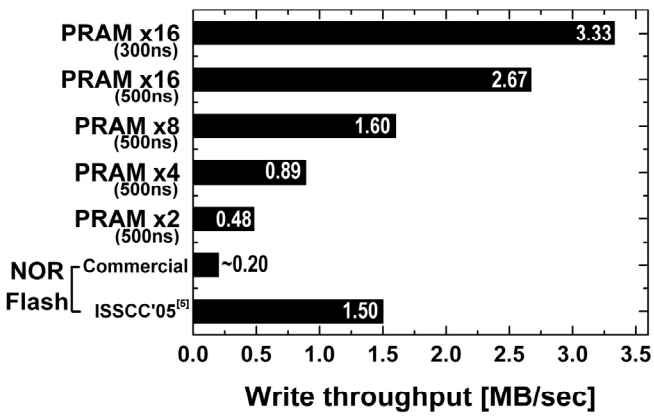

Figure 7.5.6: The write throughputs of the designed PRAM with various number of cell-parallel-write.

Continued on Page 644

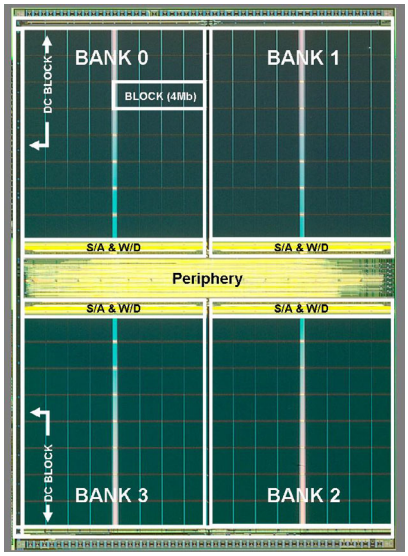

Figure 7.5.7: Micrograph of the fabricated 256Mb PRAM chip.

